

# Asymptotic Properties of One-Bit Distributed Detection with Ordered Transmissions

Paolo Braca, Stefano Marano, and Vincenzo Matta

**Abstract**—Consider a sensor network made of remote nodes connected to a common fusion center. In a recent work Blum and Sadler [1] propose the idea of ordered transmissions —sensors with more informative samples deliver their messages first— and prove that optimal detection performance can be achieved using only a subset of the total messages. Taking to one extreme this approach, we show that just a single delivering allows making the detection errors as small as desired, for a sufficiently large network size: a one-bit detection scheme can be asymptotically consistent.

The transmission ordering is based on the modulus of some local statistic (MO system). We derive analytical results proving the asymptotic consistency and, for the particular case that the local statistic is the log-likelihood ( $\ell$ -MO system), we also obtain a bound on the error convergence rate. All the theorems are proved under the general setup of random number of sensors. Computer experiments corroborate the analysis and address typical examples of applications including: non-homogeneous Poisson-deployed networks, detection by per-sensor censoring, monitoring of energy-constrained phenomenon.

## I. INTRODUCTION

**F**OLLOWING a general trend in the area of signal processing (see e.g., [2], [3]), in the last few decades there has been a considerable interest in distributed detection systems where a multitude of small sensors, properly networked to operate as a whole, takes the place of a single complex device typical of the classical system architecture.

There are several advantages that the distributed schemes boast about, in comparison with their centralized counterpart. These include robustness, scalability, flexibility, portability, failure resilience, and so forth. But the advent of distributed systems also poses new challenges to the signal processing community, since the detection layer is interleaved with the communication one so that new design trade-offs arise, yielding novel design guidelines and approaches [4]–[7].

One aspect that is of primary relevance in the implementation of many distributed detection systems is the limitation of the sensors' energy, with consequence on the sensors' capability of sensing, processing and delivering data. While a precise evaluation of the relative impact of these tasks strongly depends upon the specific network, for many wireless systems the task of communication is by far the more energy consuming [8], [9]. As a consequence, one important issue in a wireless sensor network (WSN) is how to design the system in order to reduce the amount of communication, given a desired level of detection performance.

As regards to this aspect, in the literature several approaches have been proposed. They include the design of energy efficient routing in ad-hoc networks [4], the implementation of proper strategies for the access to the common communication medium [4], and the use of censoring pioneered in [10] and further developed in many successive works, see e.g., [11]–[13]. Censoring refers to the idea of quantifying, at the sensor level, the informativeness of the sensed samples before sharing the observations with other nodes or with a sink unit: only the data that are believed informative enough are shared, while the node remains silent otherwise, optimizing the battery life.

In this work we consider a modification of this idea: we design a distributed detection system for binary hypothesis test in which the remote nodes quantify the informativeness (for detection) of their observations, and communicate their local decisions to the system, with more informative samples that are communicated first. As soon as one local decision is communicated, that is taken as the global decision of the network and the detection task terminates. Thus, censoring is obtained on a time-selective basis.

Note that two different stages exists: the sensing stage in which all the nodes observe the state of the nature to be decided over, and the successive data fusion stage in which each node is ready to deliver its local decision after a time interval that is inversely proportional to the informativeness of the observations. As the first (more informative) local decision is sent, the whole detection task is terminated with final decision equal to the local quickest decision. To make a decision, our scheme prescribes just one communication event.

A WSN can be organized according to many different architectures. One possibility is that a common fusion center (FC) exists with the role of collecting the data that the remote nodes deliver, usually after some local pre-processing. In this case, either dedicate links connecting each node to the FC exist, or there is a common channel that the sensors access by some suitable multiple access scheme.

Another common architecture lacks any central unit, and the local decisions are propagated within the network by inter-sensor communication protocols, while collaborative signal processing procedures are employed to mimic the presence of a FC. In these “fully flat” WSNs, the final decision is taken by the network in a distributed fashion and is usually shared by all the nodes, as in the consensus schemes [14]–[18].

For most part of this paper, we do not refer to any specific architecture since our goal is to investigate the detection performance of the statistics computed at the faster *firing* node of the network, which is largely independent of the specific WSN architecture. For concreteness, one can imagine that, if a FC exists, after receiving the first delivering from some

The authors are with the Department of Electronic and Computer Engineering (DIEI), University of Salerno, via Ponte don Melillo I-84084, Fisciano (SA), Italy. E-mails: {pbraca, marano, vmatta}@unisa.it.

sensor, such FC broadcasts a stopping message to all the other nodes. Conversely, in fully flat architectures, the halting command should be propagated, along with the decision, from the deciding node to all the other nodes by means of some suitable multi-hop protocol or by means of consensus algorithms. We refer to our scheme as a one-bit distributed detection; it should be noted, however, that we are disregarding the network messages required for the halting procedure.

#### A. Related work & motivations

The general approach of quickly computing an efficient detection statistic resembles the quickest detection problem originally introduced by [19]; see [20]–[22] for more recent references. The substantial difference is that quickest detection procedures are usually used to monitor in a continuous way an underlying phenomenon to discover a change in the statistics of the observed process. Conversely, in the system we design, the sensing stage and the detection stage are separated. First, the environment is sensed by all the sensors, and then a detection step is initiated. The quickness is only a property of this second stage, it is not related to event detections but it is a mean to save system energy, and there is no change in the process statistics.

More relevant to our setup is the work by Blum and Sadler [1]. They study a WSN engaged in a detection problem and conceive a multiple access architecture in which each sensor accesses the channel after a delay inversely proportional to the informativeness of its measurement. They show that, upon receiving at the FC a certain fraction of the overall sensor measurements, a decision can be made at the same performance level achievable by using the complete set of data collected by the sensors. The channel access rule, indeed, is such that the more informative samples are delivered first so that transmissions can be saved, without degradation in error probability.

In applications where sensors are severely battery-limited and relatively tiny and cheap, one can lead this approach to one extreme. What if, rather than collecting at the FC a certain fraction of the sensors' deliverings, just one single sample is considered? This would imply a significant energy saving, payed in the coin of a performance degradation with respect to the approach of [1]. However, given that sensors are tiny and cheap, performances can be improved by increasing the number of sensors, if some form of asymptotic consistency holds. This work elaborates on this concept.

#### B. Main results & organization

One main theoretical result of this paper is the proof of the asymptotic consistency of the described one-bit system when the informativeness of sensors' samples is evaluated according to the modulus  $|T(\cdot)|$  of some suitable transformations  $T(\cdot)$  of the observed samples. For instance, in the case that  $T(\cdot)$  is the identity, the idea is that extreme values of the measurements carry more information for detection with respect to "near the mean" observations.

We also consider as index of informativeness the modulus of the log-likelihood of the observed sample, which is motivated

by known results on censoring [10]. When remote sensors compute the log-likelihoods, and the delivering time is accordingly set, beside proving the asymptotic consistency of the test we derive bounds to the asymptotic rate of convergence of the error probabilities. We consider also networks whose size is random and possibly depends upon the observed data. This allows to consider very general applicative scenarios, examples of which are given in Sect. IV.

The remainder of this paper is organized as follows. The problem statement is described in Sect. II, the main results are presented in Sect. III, examples of applications are provided in Sect. IV, while in Sect. V we summarize.

## II. PROBLEM STATEMENT

### A. Preliminaries

Consider a WSN made of  $n$  remote units that sense the surrounding environment to decide which of two mutually exclusive states of the nature,  $\mathcal{H}_0$  or  $\mathcal{H}_1$ , is actually in force. The observation made by sensor  $i$  is modeled as a random variable  $X_i$ , where  $i = 1, 2, \dots, n$ , and the  $X_i$ 's are independent and identically distributed (iid) samples drawn from one of the two possible marginal probability density functions (or pdf's)  $f_X(x; \mathcal{H}_j)$ ,  $j = 0, 1$ . This simple hypothesis test can be schematically formalized as

$$\mathcal{H}_0 : X_i \sim f_X(x; \mathcal{H}_0), \quad \text{vs.} \quad \mathcal{H}_1 : X_i \sim f_X(x; \mathcal{H}_1). \quad (1)$$

We assume throughout this work that the involved random variables, taking values in  $\mathfrak{R}$ , admit densities and these densities have unbounded support, in the sense that  $\sup_x \{f_X(x) > 0\} = \infty$  and  $\inf_x \{f_X(x) > 0\} = -\infty$ .

Suppose that sensor  $i$  computes a suitable local detection statistic  $T(X_i)$ , to be compared with a certain threshold value<sup>1</sup>. If the threshold is crossed a local decision in favor of  $\mathcal{H}_1$  is made, while the local decision is for  $\mathcal{H}_0$  otherwise; let  $D_i = 0, 1$  be such decision. As proposed in [1], sensor  $i$  is programmed to communicate with the network after a time interval proportional to  $1/|T(X_i)|$ , however in this work it is supposed that sensor  $i$  delivers his local decision  $D_i$  instead of the local statistic  $T(X_i)$  as in [1]. We assume that the sensors are perfectly synchronized so that they share the same time reference. Then, the larger is  $|T(X_i)|$ , the faster is the delivering of  $D_i$  and, different from [1], in our scheme the "winner takes all". Otherwise stated, as soon as the quickest sensor delivers its own decision (say, the sensor "fires"), such decision is immediately taken as the final one for the whole system, all other transmissions by the remaining  $n - 1$  sensors are instantaneously inhibited, and the overall detection process is terminated.

While many other forms of ordering are certainly conceivable, the choice of modulus ordering leads to analytical tractability and has a precise rationale, as detailed later. Two obvious choices for the transformation  $T(\cdot)$  are the identity  $T(x) = x$ , and that based on the log-likelihood ratio

$$T(x) = L(x) := \log \frac{f_X(x; \mathcal{H}_1)}{f_X(x; \mathcal{H}_0)}.$$

<sup>1</sup>Needless to say, a likelihood ratio test would be the best. However, this might not be available, e.g., in fully or partially nonparametric setups, such that it is of interest to study general detection statistics, see also [23].

In the former case the firing time of the generic sensor  $i$  is proportional to  $1/|X_i|$ , in the latter it is proportional to  $1/|L(X_i)|$ .

The idea of accessing the channel by ordering is borrowed by [1], and the main aim of this paper is to investigate the asymptotic properties of the above distributed detector, with respect to the number  $n$  of sensors. A number of simplifying assumptions are made, including the possibility of instantaneously communicate the first local decision to sleep down the system, and the assumption of perfect time synchronism among sensors. While we use this setup to get clean analytical results and useful insights, some of the effects related to time errors and uncertainty are briefly investigated in Sect. IV-E, exhibiting a certain robustness of the proposed strategy.

### B. Detector design

First, let us specify the local testing rule of (1) at sensor  $i$

$$T(X_i) \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{>}} \gamma_n, \quad (2)$$

where  $\gamma_n$  is the detection threshold (which is allowed to depend on  $n$ ), and the local decision  $D_i$  is accordingly defined.

The transmission policies considered in this work is defined as follows.

**DEFINITION 1 (Transmission policy)** *The transmission of the local decision  $D_i$  made by the generic sensor  $i$  is activated at a time inversely proportional to the absolute value of its transformed measurement  $|T(X_i)|$ ; we call this policy MO (modulus ordered). Within the class of MO, if  $T(x) = L(x)$  the system is called  $\ell$ -MO (log-likelihood modulus ordered).  $\diamond$*

Thus, the transmission policy is identified by the transformation  $T(\cdot)$ , leading to the definition of the random variable  $Z_i := T(X_i)$ , with cumulative distribution function (cdf)  $F_Z(x; \mathcal{H}_j)$  and pdf  $f_Z(x; \mathcal{H}_j)$ , under hypothesis  $\mathcal{H}_j$  with  $j = 0, 1$ . The modulus ordering can be defined in terms of the index permutation  $\pi(\cdot)$  defined by the property that

$$|Z_{\pi(1)}| \leq |Z_{\pi(2)}| \leq \dots \leq |Z_{\pi(n)}|, \quad (3)$$

and the decision statistic of our system is  $\mathcal{M}_n := Z_{\pi(n)}$ . Therefore, the decision rule of the test (1) for the whole network is:

$$\mathcal{M}_n \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{>}} \gamma_n. \quad (4)$$

Next, consider the following extreme value statistics ( $k$  is a positive integer)

$$\mathcal{M}_n^+ := \max_{k \leq n} Z_k, \quad \mathcal{M}_n^- := -\min_{k \leq n} Z_k = \max_{k \leq n} (-Z_k),$$

whence the decision statistic in (4) can be expressed as

$$\mathcal{M}_n = \begin{cases} \mathcal{M}_n^+ & \text{if } \mathcal{M}_n^+ \geq \mathcal{M}_n^- \\ -\mathcal{M}_n^- & \text{if } \mathcal{M}_n^+ < \mathcal{M}_n^- \end{cases}. \quad (5)$$

Also, let us denote by  $F_{\mathcal{M}_n}(x; \mathcal{H}_j)$ ,  $F_{\mathcal{M}_n^+}(x; \mathcal{H}_j)$  and  $F_{\mathcal{M}_n^-}(x; \mathcal{H}_j)$  the cdf's of the above quantities under hypothesis  $\mathcal{H}_j$ , and by  $f_{\mathcal{M}_n}(x; \mathcal{H}_j)$ ,  $f_{\mathcal{M}_n^+}(x; \mathcal{H}_j)$  and  $f_{\mathcal{M}_n^-}(x; \mathcal{H}_j)$

the corresponding pdf's. Standard results of order statistics theory allows us to compute these functions as follows [24]:

$$f_{\mathcal{M}_n^+}(x; \mathcal{H}_j) = n F_Z^{n-1}(x; \mathcal{H}_j) f_Z(x; \mathcal{H}_j) \quad (6)$$

$$f_{\mathcal{M}_n^-}(x; \mathcal{H}_j) = n (1 - F_Z(-x; \mathcal{H}_j))^{n-1} f_Z(-x; \mathcal{H}_j). \quad (7)$$

As regard to the statistical distribution of  $\mathcal{M}_n$ , exploiting the results provided in [25], [26] we have

$$f_{\mathcal{M}_n}(x; \mathcal{H}_j) = n h_j^{n-1}(x) f_Z(x; \mathcal{H}_j) \quad (8)$$

where

$$h_j(x) = F_Z(|x|; \mathcal{H}_j) - F_Z(-|x|; \mathcal{H}_j). \quad (9)$$

### III. ASYMPTOTIC ANALYSIS

We are now ready to introduce the considered asymptotic setup for order statistics. We are primarily interested in the regime of large number of sensors, that is,  $n \rightarrow \infty$ . However, in many WSN applications, the number of effective sensors that contribute to the final inference is uncertain, due to several practical issues, such as failures, time-varying topologies, impaired communication, compromised nodes, and so on. Accordingly in this paper we consider the more general case of a random number of sensors; see, e.g., [27], for a discussion on the relevance of this scenario in distributed detection problems.

Formally, let  $N$  be the random number of sensors, whose distribution depends on an integer parameter<sup>2</sup>  $\nu$ . Depending on the application,  $\nu$  may represent the total number of available sensors ( $N$  of which are in fact activated), the (integer part of the) expected number of sensors  $\nu = \lfloor \mathbb{E}(N) \rfloor$ , and so forth. In order to define a proper asymptotic setup, the precise sense in which the random  $N$  diverges must be defined. A general and convenient formalization is to assume that, as the parameter  $\nu$  goes to infinity

$$\frac{N}{\nu} \rightarrow R \quad \text{in probability,} \quad (10)$$

where  $R$  is a positive random variable, see [28].

#### A. Relevant EVT background

Before illustrating the main asymptotic theorems, we briefly summarize some relevant facts from the classical literature. Let  $Y_1, Y_2, \dots, Y_n$  be a collection of iid random variables with unbounded support, and let  $M_n = \max_{k \leq n} Y_k$ .

**LEMMA 1 (Attraction)** *Under mild regularity conditions, there exist sequences of normalizing constants  $a_n, b_n$  such that*

$$\lim_{n \rightarrow \infty} F_{M_n}(a_n x + b_n) = G(x) \quad (11)$$

where  $G(x)$  is either the Gumbel distribution or the Fréchet distribution.  $\diamond$

The technical regular conditions can be found in any textbook on EVT (e.g., [29]) and the relevant features of the quoted

<sup>2</sup>Here we formally consider  $\nu$  as an integer parameter, thus the  $\nu^{th}$  element of a generic sequence  $\{a_n\}_{n=1}^{\infty}$  can be denoted as  $a_{\nu}$ . However, all the asymptotic results of this work also hold when  $\nu$  is real, and  $a_{\nu}$  is replaced by  $a_{\lfloor \nu \rfloor}$ .

distributions are as follows [29]:<sup>3</sup>

GUMBEL

$$G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty,$$

$$b_n = F_Y^{-1}\left(1 - \frac{1}{n}\right), \quad a_n = \frac{1}{n f_Y(b_n)},$$

$$\lim_{n \rightarrow \infty} (b_n/a_n) = \infty.$$

FRÉCHET

$$G(x) = \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-\xi}) & x > 0, \end{cases} \quad \xi > 0,$$

$$b_n = 0, \quad a_n = F_Y^{-1}\left(1 - \frac{1}{n}\right),$$

$$\lim_{n \rightarrow \infty} a_n = \infty.$$

An extension of Lemma 1 to the case of random number of variables has been proved by Galambos [28]:

LEMMA 2 (Attraction with random number of variables) *Let  $N$  be an integer random variable and let, as  $\nu \rightarrow \infty$ ,  $N/\nu$  converge in probability to a positive random variable  $R$ . If*

$$\lim_{n \rightarrow \infty} F_{M_n}(a_n x + b_n) = G(x), \quad (12)$$

then

$$\lim_{\nu \rightarrow \infty} F_{M_N}(a_\nu x + b_\nu) = \mathbb{E}(G^R(x)) \quad (13)$$

where the expectation is taken under the distribution of  $R$ .  $\diamond$

DEFINITION 2 (Right/Left tail dominance) *Given a random variable  $Y$ , consider the limit*

$$\lim_{x \rightarrow \infty} \frac{1 - F_Y(x)}{F_Y(-x)} \quad (14)$$

We say that  $Y$  is right-tail dominant if the above limit is  $+\infty$ , and we say that  $Y$  is left-tail dominant if the limit is zero.  $\diamond$

### B. Detection asymptotic properties

Let us explain the rationale of the modulus ordering, and for sake of simplicity assume the case in which the number of sensor is deterministic  $N = n$ . Suppose  $n$  “very large”, local decision of the firing sensor (4) is based either on the largest  $\mathcal{M}_n^+$  or on the smallest  $-\mathcal{M}_n^-$  of the  $n$  transformed samples collected by the system. However, if the right tail of the distribution  $f_Z(x; \mathcal{H}_1)$  dominates over the left tail (i.e., the right tail is heavier, or decreases slower) then with high probability the local decision is made using the largest transformed sample collected by the network. The converse happens with left-tail dominant distribution. See Fig. 1 for an instance of this effect.

In practical problems it is often the case that  $f_Z(z; \mathcal{H}_0)$  is left dominant while  $f_Z(z; \mathcal{H}_1)$  is right dominant (or viceversa), so that the hypothesis test can be thought as one comparing a very large positive sample against a very small negative value. Based on this argument, one expects that the error probability can be made smaller and smaller as  $n$  grows. Here below this intuition is verified and the sense in which the errors can be controlled is made precise.

<sup>3</sup>To avoid confusion, notice that the assumption of unbounded support rules out convergence to the third class of attraction, namely the Weibull distribution.

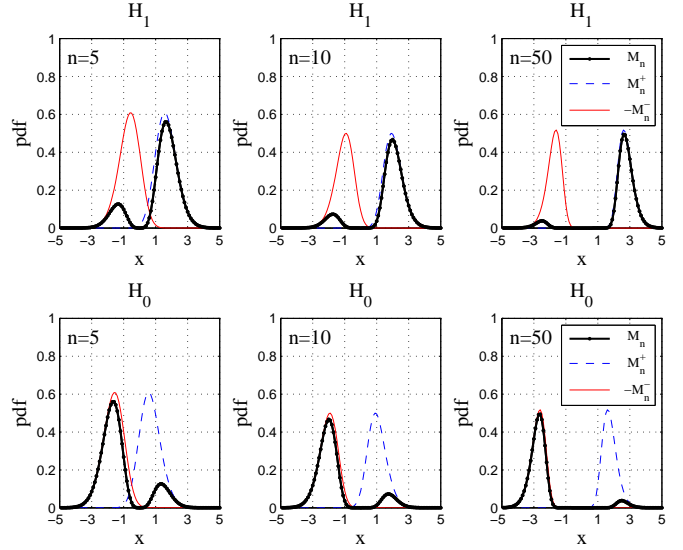


Fig. 1. Illustrative sketch of the intuition behind this work: Under  $\mathcal{H}_1$ , the pdf of  $\mathcal{M}_n$  is closer and closer to that of  $\mathcal{M}_n^+$ , as  $n$  grows, and the small peak located around the peak of  $-\mathcal{M}_n^-$  tends to vanish. Conversely, under  $\mathcal{H}_0$ , the pdf of  $\mathcal{M}_n$  approaches that of  $-\mathcal{M}_n^-$ .

THEOREM 1 (Asymptotics of MO detection statistic) *Consider an MO network with a random number of active sensors  $N$ . Suppose that*

- i) *The random variables  $Z_1, Z_2, \dots$  have unbounded support.*
- ii)  *$\mathcal{M}_n^+$  is attracted under  $\mathcal{H}_1$  with normalizing constants  $a_n^+$  and  $b_n^+$ , and limiting distribution  $G^+(x)$ . Similarly  $\mathcal{M}_n^-$  is attracted under  $\mathcal{H}_0$  with normalizing constants  $a_n^-$  and  $b_n^-$ , and limiting distribution  $G^-(x)$ .*
- iii) *The ratio  $N/\nu$  converges to a positive random variable  $R_1$  under  $\mathcal{H}_1$ , and to a positive random variable  $R_0$  under  $\mathcal{H}_0$ .*
- iv) *The random variables  $Z_i$ 's are right-tail dominant under  $\mathcal{H}_1$  and left-tail dominant under  $\mathcal{H}_0$ .*

Then

$$\frac{\mathcal{M}_N}{\mathcal{M}_N^+} \rightarrow 1, \quad \mathcal{M}_N - \mathcal{M}_N^+ \rightarrow 0, \quad \text{under } \mathcal{H}_1, \quad (15)$$

and

$$\frac{\mathcal{M}_N}{\mathcal{M}_N^-} \rightarrow -1, \quad \mathcal{M}_N + \mathcal{M}_N^- \rightarrow 0, \quad \text{under } \mathcal{H}_0, \quad (16)$$

all the convergences being in probability.  $\diamond$

*Proof:* The proof is deferred to Appendix A.

We want to stress that conditions i), ii) and iv) are by no means restrictive and hold true in a large number of practical applications. Condition iii) is a convenient way to handle with networks of random size and encompasses as special case the scenario of nonrandom  $N$ . The claim of the theorem, in words, states that the detection statistic  $\mathcal{M}_N$  tends (asymptotically) to behave like  $\mathcal{M}_N^+$  under  $\mathcal{H}_1$ , and like  $-\mathcal{M}_N^-$  under  $\mathcal{H}_0$ , see again Fig. 1. As a direct consequence of Theorem 1, we get

the following.

**COROLLARY (Attraction of the detection statistic)** *Under the same assumptions of Theorem 1*

$$\lim_{\nu \rightarrow \infty} F_{\mathcal{M}_N}(a_\nu^+ x + b_\nu^+; \mathcal{H}_1) = H^+(x), \quad (17)$$

and

$$\lim_{\nu \rightarrow \infty} F_{\mathcal{M}_N}(a_\nu^- x - b_\nu^-; \mathcal{H}_0) = H^-(x) \quad (18)$$

where

$$H^+(x) = \mathbb{E}\left(\left(G^+(x)\right)^{R_1}\right), \quad (19)$$

$$H^-(x) = 1 - \mathbb{E}\left(\left(G^-(-x)\right)^{R_0}\right). \quad (20)$$

◇

*Proof:* Assume that  $\mathcal{H}_1$  is in force. In view of Theorem 1 it follows that  $(\mathcal{M}_N - b_\nu^+)/a_\nu^+$  converges in probability to  $(\mathcal{M}_N^+ - b_\nu^+)/a_\nu^+$ . Assumption *ii*), along with a direct application of Lemma 2 implies the convergence in distribution of  $(\mathcal{M}_N^+ - b_\nu^+)/a_\nu^+$ . Then, the convergence in distribution of  $(\mathcal{M}_N - b_\nu^+)/a_\nu^+$  claimed in (17) follows by a direct application of Theorem 2.7 in [30]. The proof for  $\mathcal{H}_0$  is similar. •

As to the performance of the hypothesis test, this is expressed in terms of the false alarm and miss detection probabilities

$$\alpha_\nu = \mathbb{P}(\text{decide } \mathcal{H}_1; \mathcal{H}_0) = \mathbb{P}(\mathcal{M}_N \geq \gamma_\nu; \mathcal{H}_0), \quad (21)$$

$$\beta_\nu = \mathbb{P}(\text{decide } \mathcal{H}_0; \mathcal{H}_1) = \mathbb{P}(\mathcal{M}_N < \gamma_\nu; \mathcal{H}_1). \quad (22)$$

We note explicitly that, being  $N$  random, the threshold of the test cannot be set as a function of that, but rather it must be controlled by the parameter  $\nu$ , that is clearly assumed known in order to fix the threshold value.

We consider the classical setup where a prescribed (asymptotic) false-alarm level  $\alpha$  is imposed, while it is required that  $\beta_\nu$  vanishes with increasing  $\nu$ . In the light of Theorem 1, it is reasonable to impose an asymptotic false-alarm level  $\alpha$  based on the asymptotic distribution under  $\mathcal{H}_0$ . We indeed know that  $(\mathcal{M}_N + b_\nu^-)/a_\nu^-$  converges in distribution toward  $H^-(x)$ . This implies that the threshold  $\gamma_\nu = a_\nu^- \gamma - b_\nu^-$ , with

$$\alpha = 1 - H^-(\gamma) = \mathbb{E}\left(\left(G^-(-\gamma)\right)^{R_0}\right),$$

achieves the asymptotic false alarm  $\alpha$ .

For this computation, it is useful to define the false alarm  $\tilde{\alpha}$  corresponding to a system with deterministic number of sensors, that is

$$G^-(-\gamma) = \tilde{\alpha} \implies \gamma = \begin{cases} \log \log(1/\tilde{\alpha}) & \text{GUMBEL} \\ -(\log(1/\tilde{\alpha}))^{-\frac{1}{\xi}} & \text{FRÉCHET} \end{cases} \quad (23)$$

The required false alarm  $\alpha$  can be computed as a function of  $\tilde{\alpha}$ , by

$$\tilde{\alpha} : \mathbb{E}(\tilde{\alpha}^{R_0}) = \alpha. \quad (24)$$

The threshold  $\gamma_\nu = a_\nu^- \gamma - b_\nu^-$  is selected by using the asymptotic  $\mathcal{H}_0$ -distribution  $H^-(x)$ . An alternative might be that of imposing the strict equality  $\alpha_\nu = \alpha$  for any finite  $\nu$ . This, however, would require exact knowledge of the cdf of the detection statistic for any finite  $\nu$ , which is usually unavailable.

On the other hand, it is possible to use the asymptotic “similarity” (under  $\mathcal{H}_0$ ) between  $\mathcal{M}_N$  and  $-\mathcal{M}_N^-$  to set a new

threshold as  $(1 - F_Z(\gamma_\nu; \mathcal{H}_0))^\nu = \tilde{\alpha}$ , which can be shown to achieve asymptotically the desired false-alarm level. These results are summarized in the following theorem.

**THEOREM 2 (MO consistency)** *Under the assumptions of Theorem 1:*

*i) The nonparametric setting  $\gamma_\nu = 0$  ensures that*

$$\alpha_\nu + \beta_\nu \rightarrow 0. \quad (25)$$

*ii) Let the detection threshold be either*

$$\gamma_\nu = a_\nu^- \gamma - b_\nu^-, \quad (26)$$

*where  $\gamma$  solves  $1 - H^-(\gamma) = \alpha$ , or*

$$\gamma_\nu = F_Z^{-1}\left(1 - \tilde{\alpha}^{\frac{1}{\nu}}; \mathcal{H}_0\right), \quad (27)$$

*where  $\tilde{\alpha}$  solves  $\mathbb{E}(\tilde{\alpha}^{R_0}) = \alpha$ . Then:*

$$\alpha_\nu \rightarrow \alpha, \quad \text{and} \quad \beta_\nu \rightarrow 0. \quad (28)$$

◇

*Proof:* The proof is deferred to Appendix B.

*Remark.* The threshold setting used in eq. (25) does not require any a-priori knowledge of the statistics, that amounts to a *nonparametric* threshold setting. This may be convenient in practical applications where limited knowledge of the statistical model is available to the remote nodes.

The above theorems are valid for a general MO network. For specific detection problems and/or local transformations, more powerful results might be obtained. This is the case of an  $\ell$ -MO strategy (namely, when  $Z_i$  is the log-likelihood of the observations) applied to the shift-in-mean problems of the kind

$$f_X(x; \mathcal{H}_0) = \phi(x + \theta_0), \quad f_X(x; \mathcal{H}_1) = \phi(x - \theta_1) \quad (29)$$

where  $\phi(x)$  is an even function,  $\phi(x) > 0 \forall x$ ,  $\theta_0 \geq 0$  and  $\theta_1 > 0$ . For this scenario we prove the following

**THEOREM 3 ( $\ell$ -MO properties)** *Assume that the local log-likelihoods fulfill conditions i), ii) and iii) of Theorem 1. Then, condition iv) is automatically verified, and the results of Theorems 1 and 2 apply. In addition, if  $N$  is independent of the observations, the following upper bound on the miss detection probability holds*

$$\beta_\nu \leq e^{\gamma_\nu}. \quad (30)$$

◇

*Proof:* The proof is deferred to Appendix C.

#### IV. APPLICATIONS

To illustrate the above results, we now focus on sensor network applications. Both MO and  $\ell$ -MO systems are investigated for different case studies, with the twofold goal of providing a numerical check for the asymptotic convergence claimed in the theoretical results, and of investigating the effect of a moderately small number of sensors. We also consider networks of random size and a typical application example from the distributed detection domain, such as censoring sensor systems. Finally we address the case where the random

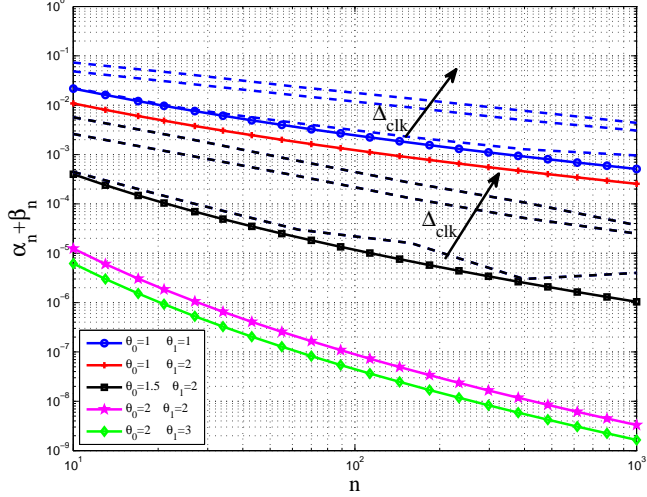


Fig. 2. MO transmission policy. Error  $\alpha_n + \beta_n$  for the Gaussian nonparametric example, with  $\sigma = 1$  and different combinations of  $\theta_0$  and  $\theta_1$ . The solid curves are obtained by numerical integration based on eq. (8). Dashed lines refer to clock offset discussed later, in Sect. IV-E: It is shown the effect of synchronism errors for  $\Delta_{clk} = 0.1, 1, 2$  with respect to both the nominal case  $\theta_0 = \theta_1 = 1$  and  $\theta_0 = 1.5, \theta_1 = 2$ .

network size depends upon the observations, and we briefly touch upon the robustness of the detection system to timing offsets.

#### A. Gaussian observations

Let us start by considering the following Gaussian observation model ( $\mathcal{N}(a, b)$  is our shortcut for a Gaussian distribution with mean  $a$  and standard deviation  $b$ ):

$$\mathcal{H}_0 : X_i \sim \mathcal{N}(-\theta_0, \sigma), \quad \text{vs.} \quad \mathcal{H}_1 : X_i \sim \mathcal{N}(\theta_1, \sigma), \quad (31)$$

where  $\theta_0, \theta_1$  and  $\sigma$  are positive parameters, and the number of sensors  $n$  is deterministic (namely, here we set  $N = \nu = n$ ). Let us consider first the MO policy with  $T(x) = x$ . According to Theorem 2,  $\alpha_n \rightarrow 0$  and  $\beta_n \rightarrow 0$ . This is true even if  $\theta_0, \theta_1, \sigma$  and  $n$ , are all unknown. In this case, we are faced with a fully nonparametric test in which the sensors have no knowledge of the parameters appearing in (31) and they accordingly use a zero threshold, see Theorem 2, part *i*). The results are shown in Fig. 2, where the corresponding error probabilities (solid curves) have been obtained by numerical integration based on expression (8): as predicted both the error probabilities go to zero, with a rate that depends upon the system parameters. The dashed curves refer to the effect of clock offsets, and we comment on this later.

With reference to the same observation model (31), suppose now that the parameters are known and that the  $\ell$ -MO policy is in order. It is clear that the  $\ell$ -MO policy cannot be implemented without the knowledge of the distribution parameters, since it requires the computation of the likelihood. We assume again that  $n$  is deterministic. This case lies in the application domain of Theorem 3 and the error probabilities, still computed by numerical integration based on (8), are illustrated in Fig. 3. It is worth noting that the asymptotic value

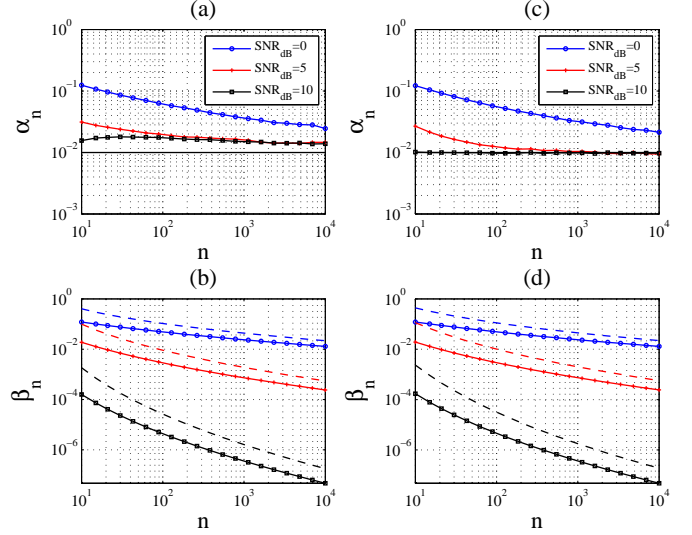


Fig. 3.  $\ell$ -MO transmission policy. Error probabilities  $\alpha_n$  and  $\beta_n$  for the Gaussian example with known parameters. The curves are obtained by numerical integration based on eq. (8), and are parametrized in  $\text{SNR} = (\theta_1 + \theta_0)/\sigma$ . Panels (a) and (b) refer to the threshold setting given in (26), while (c) and (d) refer to that in (27). In the lower plots are also shown, as dashed lines, the miss detection bounds (30).

is approached faster for larger values of  $\text{SNR} = (\theta_1 + \theta_0)/\sigma$ . This should be expected because, as it can be easily seen, the tail dominance is “stronger” when the SNR grows. Panels (a) and (b) refer to the threshold setting given in (26), while (c) and (d) refer to the threshold in (27). We see that  $\alpha_n$  converges to the desired asymptotic value (set to  $\alpha = 10^{-2}$  in the figure); however, in (c) the convergence is somehow faster than that in (a), suggesting that the threshold setting (27) provides, in this example, some advantage. As claimed in Theorem 3, we see that  $\beta_n \rightarrow 0$  as shown in panels (b) and (d). Note that the curves in (b) and (d) are very similar, which reveals that the selection of the threshold between the two alternatives, is not critical with respect to  $\beta_n$ . Also shown is the upper bound on miss detection probability given by (30), that in this case (and for both the thresholds) can be approximated by the simple expression  $\beta_n \leq \exp(-\text{SNR}\sqrt{2} \log n)$ , after neglecting terms of higher order in  $n$ .

#### B. Networks of random size

The powerfulness of the theorems presented in the previous section allows us to consider the more general setting of network of random size  $N$ . Consider hence the following scenario.

- Sensors are randomly deployed in a two-dimensional region  $\mathcal{A}$ , according to a non-homogeneous Poisson field. The intensity function of this field is  $\lambda(x)$ ,  $x \in \mathcal{A}$ , such that the average number of sensors in the region  $\mathcal{A}$  is  $\int_{x \in \mathcal{A}} \lambda(x) dx$ .
- Some sensors are impaired before (or at) the act of communication. Thus, the number of active sensors  $N$  is a subset of those globally available. The probability of a failure is unknown to the network, and is accordingly modeled as a random variable  $Q$ , independent of the deploying process.
- Conditioned on  $Q = q$ , the active sensors are selected



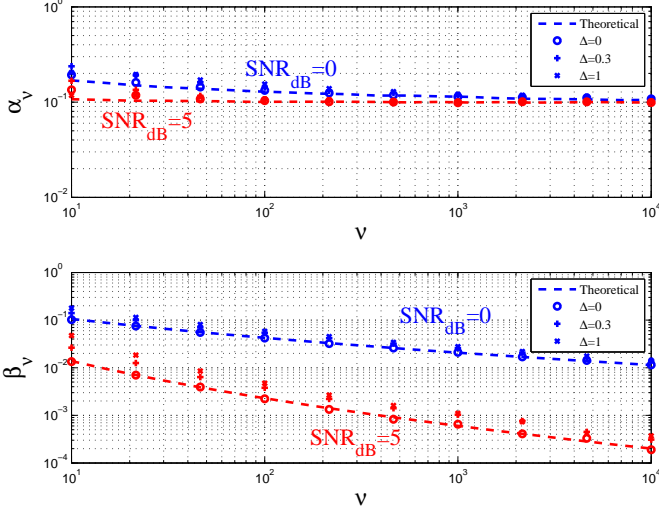


Fig. 4.  $\ell$ -MO transmission policy with networks of random size and Gaussian observation model. The values of  $\alpha_n$  and  $\beta_n$  are plotted, for different values of  $\text{SNR} = (\theta_1 + \theta_0)/\sigma$  and different values of  $\Delta$ . The curves labeled with “theoretical” are shown for comparison and refers to  $N$  deterministic, while those with  $\Delta = 0$  refer to the simple Poisson model with deterministic mean value.

independently and with probability  $q$  from the total number of sensors available in  $\mathcal{A}$ . Given  $Q = q$ , the number of active sensors becomes a Poisson random variable with mean value  $q \int_{\mathbf{x} \in \mathcal{A}} \lambda(\mathbf{x}) d\mathbf{x}$ .

- Accordingly, the average number of active sensors is

$$\mathbb{E}(N) = \mathbb{E}(Q) \int_{\mathbf{x} \in \mathcal{A}} \lambda(\mathbf{x}) d\mathbf{x} := \nu.$$

By introducing the normalized random variable  $R = Q/\mathbb{E}(Q)$ , we have

$$\mathbb{P}(N = n | R = r) = \frac{(\nu r)^n}{n!} e^{-\nu r}.$$

- We focus on the asymptotic regime of increasingly large sensors, which in this context is formalized by  $\nu \rightarrow \infty$ . From a practical perspective, note that an increasingly large value of  $\nu$  may be due to an increasing large sensor density  $\lambda(\mathbf{x})$ , that corresponds to the asymptotic regime of a *dense* network (recall however that sensors’ observations are iid), or to an increasingly large surveyed region  $\mathcal{A}$ , corresponding to the asymptotic regime of a *large* network.

- It is easy to show that<sup>4</sup>  $N/\nu \rightarrow R$  in probability, when  $\nu \rightarrow \infty$ .

As an example, let us consider again the observation model in (31) with  $\ell$ -MO transmission policy, but assume now that the effective network size  $N$  is random according to the model described above, and suppose that  $Q = \mathbb{E}(Q) + U$ , with  $U \sim \mathcal{U}(-\Delta/2, \Delta/2)$ , where  $\mathcal{U}(a, b)$  stems for the uniform distribution with support  $(a, b)$ . In Fig. 4 the false alarm and miss detection probabilities, parametrized in  $\text{SNR} = (\theta_1 + \theta_0)/\sigma$ , are shown for different values of  $\Delta$ , with  $\mathbb{E}(Q) = 0.5$ . The curves are obtained by means of Monte Carlo computer experiments,

<sup>4</sup>In fact,  $\mathbb{P}\left(\left|\frac{N}{\nu} - r\right| > \epsilon | R = r\right) \rightarrow 0$  by the weak law of large numbers. Convergence in probability of the unconditioned random variable  $N$  to  $R$  easily follows by Lebesgue dominated convergence theorem.

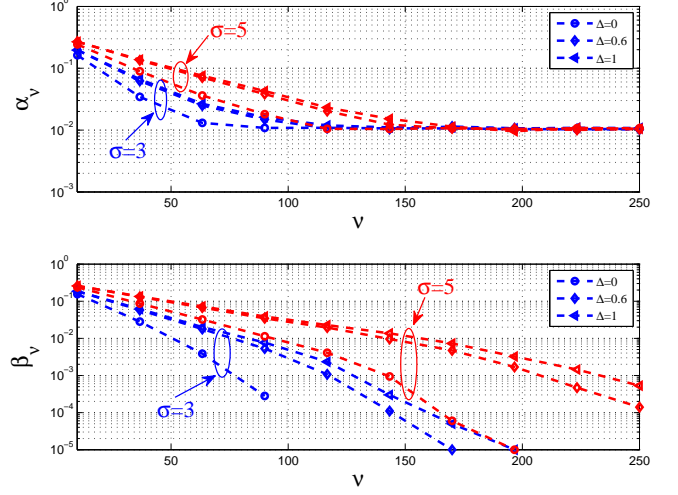


Fig. 5. Censoring transmission policy with networks of random size and non-Gaussian observation model. The values of  $\alpha_n$  and  $\beta_n$  are plotted, for two values of  $\sigma$  and different values of the randomness index  $\Delta$ . The curves with  $\Delta = 0$  refer to the simple Poisson model with deterministic mean value.

except those labeled as “theoretical”. These, plotted for comparison, refer to the case of  $N = \nu$  deterministic and are obtained by numerical integration.

We see that the false alarm probability  $\alpha_n$  converges to its limiting value  $\alpha = 10^{-1}$ , with a convergence rate that is faster for larger SNRs. The same is true for the convergence to zero of  $\beta_n$  shown in the lower plot. It is also worth noting that the limit value of  $\alpha_n$  is approached faster in the case of higher SNR, and almost at the same rate for  $N$  random and  $N$  deterministic. When the randomness grows, namely  $\Delta$  becomes larger, we see that for the miss detection probability the convergence is slightly slowed down.

### C. An example with censoring

Consider again a network of random size, as described in Sect. IV-B, but let us explore an example in which the transmission policy is based on a censoring strategy. Censoring techniques are commonly implemented in WSNs working under severe communication constraints, and amount to discard sensors’ observations considered poorly informative for the detection purpose, see, e.g., [10]–[13], [31]. This can be obtained by selecting the transformation for the transmission policy according to the censoring rule:

$$T(x) = \begin{cases} x & \text{if } |x| \geq \theta_c \\ 0 & \text{if } |x| < \theta_c \end{cases}$$

where  $\theta_c > 0$  is the censoring threshold.

To enrich the example, we adopt an observation model different from the Gaussian one considered so far. Specifically, assume that under  $\mathcal{H}_1$  the sensors observe  $X_i = W_i$  while under  $\mathcal{H}_0$  they observe  $X_i = -W_i$ . The random variables  $W_i$ ’s are iid with pdf given by a mixture between a Gaussian and a Pareto density:

$$p \frac{1}{\sqrt{2\pi}\sigma^2} e^{-x^2/2\sigma^2} + (1-p) \frac{b}{\theta} \left(\frac{x}{\theta}\right)^{-b-1} u(x-\theta)$$

where  $u(x)$  is the unit step function and  $0 < p < 1$ . In the computer experiments we set  $\theta_c = \theta = b = 1$ ,  $p = 0.5$ , and  $\mathbb{E}(Q) = 0.7$ . The detection errors  $\alpha_n$  and  $\beta_n$ , computed by Monte Carlo simulations, are displayed in Fig. 5 for two values of  $\sigma$  and different values of  $\Delta$ . The general behavior is similar to that of Fig. 4: in particular the convergence is faster when there is less randomness in the system, as quantified by the value of  $\Delta$ .

#### D. Observation-dependent network size

The previous examples demonstrate the large versatility of the theorems provided in Sect. III that ensure the asymptotic convergence of detection tests under a very broad class of applicative scenarios of practical relevance, including different transmission policies, different observation distributions, and very general network models. We now go even further by letting the random network size  $N$  to be dependent upon the sensors' observations—a possibility well encompassed in the theorems of Sect. III.

Let, as usual, the  $i^{\text{th}}$  sensor of the network monitor the physical phenomenon of interest by collecting the sample  $X_i$ . Suppose further that  $X_i = S_i + W_i$ , where the random variable  $S_i$  models the intrinsic state of the observed phenomenon, while the random variable  $W_i$  models the sensor measurement process; these two components are mutually independent and independent across sensors.

As a distinct feature of this new scenario, we assume that the number of samples collected by the system is dependent upon the nature of the observed phenomenon, in such a way that the monitoring stage is ended at a certain random sample number<sup>5</sup>:

$$N : \inf \{n : \varphi(S_1, S_2, \dots, S_n) > \nu\}, \quad \varphi(\cdot) > 0. \quad (32)$$

To fix ideas,  $\varphi(\cdot)$  might be thought as a measure of the energy emitted by the surveyed physical system, which is assumed to be limited. In our asymptotic framework, we are interested in increasingly large values of  $N$ , and this explains why the threshold in (32) has been just set to  $\nu$ .

Consider for example a Gaussian shift-in-mean problem, with  $S_i \sim \mathcal{N}(-\theta_0, \sigma_s)$  under  $\mathcal{H}_0$  and  $S_i \sim \mathcal{N}(\theta_1, \sigma_s)$  under  $\mathcal{H}_1$ , and with  $W_i \sim \mathcal{N}(0, \sigma_w)$  under both hypotheses. Assume also that the stopping rule for the acquisition process is  $N : \inf \{n : \sum_{i=1}^n S_i^2 > \nu\}$ , where  $\varphi(s_1, s_2, \dots, s_n) = \sum_{i=1}^n s_i^2$  quantifies an energy expense (but for a normalization factor). By the theory of renewal processes [32], we know, for  $j = 0, 1$

$$\frac{N}{\nu} \rightarrow \frac{1}{\theta_j^2 + \sigma_s^2} \quad \text{under } \mathcal{H}_j, \quad (33)$$

where the limit is to be intended with probability one. Note further that the parameter  $\nu$  is still related to the expected number of sensors. Indeed we also have [32], for  $j = 0, 1$

$$\frac{\mathbb{E}(N)}{\nu} \rightarrow \frac{1}{\theta_j^2 + \sigma_s^2} \quad \text{under } \mathcal{H}_j. \quad (34)$$

<sup>5</sup>The random “time” defined in eq. (32) is by construction a Markov time, in that the event  $N = n$  is determined by the observations of the first  $n$  samples. Moreover, it will be a stopping time, provided that  $\mathbb{P}(N < \infty) = 1$ .

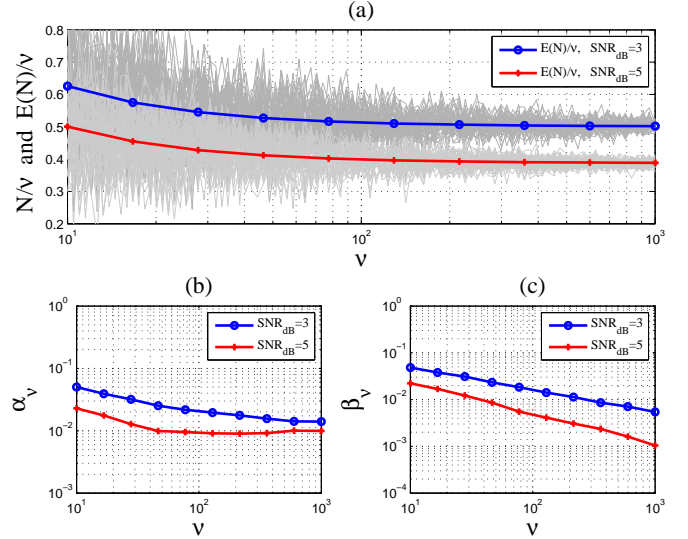


Fig. 6.  $\ell$ -MO transmission policy for a network whose size depends upon the observations. Top plot (valid for both  $\mathcal{H}_0$  and  $\mathcal{H}_1$ ) shows several realizations of  $N/\nu$  (tiny curves) to illustrate the convergence with probability one in (33), along with  $\mathbb{E}(N)/\nu$  that converges to the same limit according to (34). Lower plots show the error probabilities for an asymptotic false alarm  $\alpha = 10^{-2}$ , with threshold set by (27).

We apply the proposed one-bit detection strategy to the above situation with  $\theta_0 = \theta_1$ ,  $\sigma_s = \sigma_w = 1$  and  $\text{SNR} = (\theta_1 + \theta_0)/\sqrt{\sigma_s^2 + \sigma_w^2}$ . The pertinent results are displayed in Fig. 6, for an asymptotic false alarm probability  $\alpha = 10^{-2}$ , with threshold set by eq. (27) and using an  $\ell$ -MO policy. In Fig. 6 (a) we show the convergence of the stopping number  $N$  as given in (33) and (34). In Fig. 6 (b) and (c), we show the behavior of the error probabilities.

#### E. Resilience to clock offset

All the cases addressed above fall in the assumption of our theorems. As last example, we want to investigate briefly the robustness of the results with respect to models that slightly deviate from the formal assumptions of the theorems presented in Sect. III. In particular, we have assumed so far that sensors are perfectly synchronized, and one should note that the transmission policy strictly relies on such assumption. What if time references of the sensors are slightly misaligned? To make things simple, suppose that the clock of the generic sensor  $i$  is perturbed by  $U_i \sim \mathcal{U}(-\Delta_{clk}/2, \Delta_{clk}/2)$  that models the timing offset. In other words, sensor  $i$  will attempt to transmit its local decision at the time instant  $1/|T(X_i)| + U_i$ . Let us refer, for simplicity, to the Gaussian shift-in-mean example provided in Sect. IV-A. The dashed curves in Fig. 2 show the effect of timing errors with different  $\Delta_{clk} = 0.1, 1, 2$ , with respect to the nominal cases  $\theta_0 = \theta_1 = 1$  and  $\theta_0 = 1.5, \theta_1 = 2$ .

As it can be seen, the test consistency seems to be preserved, thus evidencing a certain robustness of the proposed strategy. On the other hand, and perhaps unsurprisingly, by increasing the offset error  $\Delta_{clk}$ , the performance worsen in the sense that the rate of convergence is slower, as consequence of the fact that the firing sensor may be different from the largest in modulus which conveys the largest information. Quantifying



the effect of  $\Delta_{clk}$  on the convergence rate and understanding whether a certain  $\Delta_{clk}$  exists such that asymptotic convergence of the errors is lost, remain open problems.

## V. SUMMARY

Distributed detection in large wireless sensor networks can be performed by the transmission of a single bit, exploiting the idea of ordered transmission policies. After casting such problem in a precise mathematical framework, we propose an easy-to-implement distributed statistical test whose asymptotic consistency is formally proved: Both the error probabilities can be controlled in the asymptotic regime of large network size, under a very broad class of observation models — from classical Gaussian shift-in-mean to fairly more general measurement settings— and applicative domains, including: nonparametric tests, likelihood-based transmission policies, censored systems, random network size, and observation-dependent sensor number.

## APPENDIX A PROOF OF THEOREM 1

We shall work under  $\mathcal{H}_1$ , thus proving eqs. (15), and consistently skip the explicit dependence upon the hypothesis for notational ease. The proof of eqs. (16) follows straightforwardly. Let us introduce the sequence of events  $\mathcal{E}_\nu = \{\mathcal{M}_N^+ \geq \mathcal{M}_N^-\}$ . In view of the definition of the detection statistic  $\mathcal{M}_N$ , the claim of the theorem will be certainly true if  $\mathbb{P}(\mathcal{E}_\nu) \rightarrow 1$ . In order to show that this convergence actually takes place, let us elaborate as follows. Let  $0 < x_0 < x_1 < +\infty$  such that

$$\mathbb{P}(R \in [x_0, x_1]) \geq 1 - \epsilon, \quad (35)$$

where, we recall,  $N/\nu$  converges in probability to  $R$ . We are now legitimate to write

$$\begin{aligned} \mathbb{P}(\mathcal{E}_\nu) &\geq \mathbb{P}\left(\mathcal{M}_N^+ \geq \mathcal{M}_N^-, \frac{N}{\nu} \in [x_0, x_1]\right) \\ &\geq \mathbb{P}\left(\mathcal{M}_{n_0}^+ \geq \mathcal{M}_{n_1}^-, \frac{N}{\nu} \in [x_0, x_1]\right) \end{aligned} \quad (36)$$

where we have defined  $n_j = \lfloor \nu x_j \rfloor$ ,  $j = 0, 1$ , and the last inequality follows by obvious properties of maxima and minima. Assume for now that

$$\lim_{\nu \rightarrow \infty} \mathbb{P}(\mathcal{M}_{n_0}^+ \geq \mathcal{M}_{n_1}^-) = 1. \quad (37)$$

This would imply that the last limit in eq. (36) equals

$$\lim_{\nu \rightarrow \infty} \mathbb{P}\left(\frac{N}{\nu} \in [x_0, x_1]\right) = \mathbb{P}(R \in [x_0, x_1]) \geq 1 - \epsilon$$

where the inequality follows by eq. (35). Inequality (36),  $\epsilon$  being arbitrary, implies  $\liminf_{\nu \rightarrow \infty} \mathbb{P}(\mathcal{E}_\nu) = 1$ , and hence  $\mathbb{P}(\mathcal{E}_\nu) \rightarrow 1$  as  $\nu$  diverges.

It remains thus to show that eq. (37) holds. To this aim, it is expedient to work in terms of the normalized variables  $\widetilde{\mathcal{M}}_{n_0}^+ = (\mathcal{M}_{n_0}^+ - b_{n_1}^+)/a_{n_1}^+$  and  $\widetilde{\mathcal{M}}_{n_1}^- = (\mathcal{M}_{n_1}^- - b_{n_1}^+)/a_{n_1}^+$ . Note first that, by assumption *ii*), it is straightforward to conclude that

$$\lim_{\nu \rightarrow \infty} \mathbb{P}(\widetilde{\mathcal{M}}_{n_0}^+ \leq x) = (G^+(x))^\eta, \quad \text{where } \eta = \frac{x_0}{x_1}. \quad (38)$$

Furthermore, by assumption *iv*) we know that the  $Z_i$ 's are right-tail dominant under  $\mathcal{H}_1$ , which implies (see [33], proof of Theorem 2.1), for all  $x$  with  $G^+(x) \neq 0$  and  $G^+(x) \neq 1$ , that

$$\mathbb{P}(\widetilde{\mathcal{M}}_{n_1}^- < x) \rightarrow 1. \quad (39)$$

It is convenient to study separately the different admissible attraction domains. Let us first consider the case that  $G^+(x)$  is a Gumbel distribution. We have

$$\begin{aligned} \mathbb{P}(\widetilde{\mathcal{M}}_{n_0}^+ \geq \widetilde{\mathcal{M}}_{n_1}^-) &\geq \mathbb{P}(\widetilde{\mathcal{M}}_{n_0}^+ - \widetilde{\mathcal{M}}_{n_1}^- \geq 0, \widetilde{\mathcal{M}}_{n_1}^- < -\delta) \\ &\geq \mathbb{P}(\widetilde{\mathcal{M}}_{n_0}^+ + \delta \geq 0, \widetilde{\mathcal{M}}_{n_1}^- < -\delta) \end{aligned} \quad (40)$$

where  $\delta > 0$  is arbitrarily large. On the other hand, eq. (39), along with eq. (38) implies

$$\lim_{\nu \rightarrow \infty} \mathbb{P}(\widetilde{\mathcal{M}}_{n_0}^+ + \delta \geq 0, \widetilde{\mathcal{M}}_{n_1}^- < -\delta) = 1 - e^{-\eta e^\delta}, \quad (41)$$

yielding, in the light of eq. (40) and being  $\delta$  arbitrary,

$$\liminf_{\nu \rightarrow \infty} \mathbb{P}(\widetilde{\mathcal{M}}_{n_0}^+ \geq \widetilde{\mathcal{M}}_{n_1}^-) = 1.$$

Let us switch now to the case that  $G^+(x)$  is Fréchet distributed. We first note that  $\widetilde{\mathcal{M}}_{n_1}^-$  now vanishes in probability. Indeed, thanks to eq. (39), we have  $\mathbb{P}(\widetilde{\mathcal{M}}_{n_1}^- > \epsilon) \rightarrow 0$  and

$$\mathbb{P}(\widetilde{\mathcal{M}}_{n_1}^- < -\epsilon) \leq \mathbb{P}(\widetilde{\mathcal{M}}_{n_1}^- < 0) = (1 - F_Z(0))^{n_1} \rightarrow 0,$$

having used the fact that, for the Fréchet domain of attraction,  $b_n^+ = 0$ . Moreover, by eq. (38), the sequence  $\widetilde{\mathcal{M}}_{n_0}^+$  converges to a Fréchet random variable. Slutsky's theorem allows to conclude that the sequence  $\widetilde{\mathcal{M}}_{n_0}^+ - \widetilde{\mathcal{M}}_{n_1}^-$  converges in distribution to a non-negative random variable, implying the desired result (37). •

## APPENDIX B PROOF OF THEOREM 2

Let's start with part *i*), and accordingly consider the term

$$\beta_\nu = \mathbb{P}(\mathcal{M}_N < 0; \mathcal{H}_1). \quad (42)$$

Now, for the case that  $G^+(x)$  is Gumbel, by the convergence in distribution of  $(\mathcal{M}_N - b_\nu^+)/a_\nu^+$ , and the divergence of the term  $b_\nu^+/a_\nu^+$ , we desume that  $\mathcal{M}_N/b_\nu^+ \rightarrow 1$  in probability, implying  $\beta_\nu \rightarrow 0$  in the light of eq. (42). For the case that  $G^+(x)$  is Fréchet, we know that  $\mathcal{M}_N/a_\nu^+$  converges in distribution to  $H^+(x)$ , which is supported on  $x > 0$ , and again  $\beta_\nu \rightarrow 0$ . Similar reasoning will lead to  $\alpha_\nu \rightarrow 0$ .

Let us switch to the part *ii*), and consider first  $\gamma_\nu$  as in eq. (26). For this case convergence of  $\alpha_\nu$  toward  $\alpha$  is nothing but eq. (18). Let us move to  $\gamma_\nu$  defined as in eq. (27). The attraction properties of  $\mathcal{M}_n^-$  imply

$$\mathbb{P}\left(\frac{\mathcal{M}_n^- - b_n^-}{a_n^-} < -\gamma; \mathcal{H}_0\right) \rightarrow G^-(-\gamma) = \tilde{\alpha}$$

where the last equality follows by eq. (23). On the other hand, by the definition of the refined threshold  $\gamma_n$  in eq. (27)

$$\mathbb{P}\left(\frac{\mathcal{M}_n^- - b_n^-}{a_n^-} < -\frac{\gamma_n + b_n^-}{a_n^-}; \mathcal{H}_0\right) = \tilde{\alpha}.$$

In view of Lemma 11.2.1 in [34], this allows concluding that  $(\gamma_n + b_n^-)/a_n^- \rightarrow \gamma$ . By eq. (18) we have thus

$$\begin{aligned}\alpha_\nu &= \mathbb{P}(\mathcal{M}_N > \gamma_\nu; \mathcal{H}_0) \\ &= \mathbb{P}\left(\frac{\mathcal{M}_N + b_\nu^-}{a_\nu^-} > \frac{\gamma_\nu + b_\nu^-}{a_\nu^-}; \mathcal{H}_0\right) \rightarrow \mathbb{E}(\tilde{\alpha}^{R_0}) = \alpha,\end{aligned}$$

the last equality following by eq. (24).

Let us now switch to the analysis of  $\beta_\nu$ . Note that the threshold  $\gamma_\nu$  in eq. (26) is negative, at least for sufficiently large  $\nu$ . Indeed, if  $G^-(x)$  is Gumbel,  $b_\nu^-/a_\nu^- \rightarrow \infty$  and  $a_\nu^- > 0$ ; if  $G^-(x)$  is Fréchet,  $\gamma < 0$ , being the support of  $1 - H^-(x)$  confined to the negative axis, see eq. (20). Simple inspection show that the threshold  $\gamma_\nu$  in eq. (27) is as well negative, at least for sufficiently large  $\nu$ . Thus, for sufficiently large  $\nu$  and for both choices of the thresholds one can write  $\beta_\nu \leq \mathbb{P}(\mathcal{M}_N < 0; \mathcal{H}_1)$ , and the proof is now complete. •

## APPENDIX C PROOF OF THEOREM 3

Let us first check the validity of condition *iv*) in Theorem 1. The symmetry of the function  $\phi(x)$  imply, for the considered shift-in-mean problem,  $f_Z(x; \mathcal{H}_1) = f_Z(-x; \mathcal{H}_0)$ . The well-known nesting rule for the log-likelihoods [35] further gives  $\log \frac{f_Z(x; \mathcal{H}_1)}{f_Z(x; \mathcal{H}_0)} = x$ . Combining the above results gives  $f_Z(x; \mathcal{H}_0) = f_Z(-x; \mathcal{H}_0)e^{-x}$ , and  $f_Z(x; \mathcal{H}_1) = f_Z(-x; \mathcal{H}_1)e^x$ , which clearly implies that  $Z$  is right-tail dominant under  $\mathcal{H}_1$  and left-tail dominant under  $\mathcal{H}_0$ .

Let us now prove eq. (30). To this aim, we write the log-likelihood ratio of  $\mathcal{M}_N$

$$\begin{aligned}\log \frac{\sum_{n=1}^{\infty} f_{\mathcal{M}_n}(x; \mathcal{H}_1) \mathbb{P}(N = n)}{\sum_{n=1}^{\infty} f_{\mathcal{M}_n}(x; \mathcal{H}_0) \mathbb{P}(N = n)} \\ = \log \frac{f_Z(x; \mathcal{H}_1) \sum_{n=1}^{\infty} n h_1^{n-1}(x) \mathbb{P}(N = n)}{f_Z(x; \mathcal{H}_0) \sum_{n=1}^{\infty} n h_0^{n-1}(x) \mathbb{P}(N = n)} \\ = x + \log \frac{\sum_{n=1}^{\infty} n h_1^{n-1}(x) \mathbb{P}(N = n)}{\sum_{n=1}^{\infty} n h_0^{n-1}(x) \mathbb{P}(N = n)}\end{aligned}$$

where in the last equality we again applied the nesting rule. Moreover, it is easy to check that, in the shift-in-mean case with even  $\phi(x)$ , we have  $h_1(x) = h_0(x)$ , finally yielding  $\log \frac{f_{\mathcal{M}_N}(x; \mathcal{H}_1)}{f_{\mathcal{M}_N}(x; \mathcal{H}_0)} = x$ . At this point we are legitimate to use the Chernoff bound:  $\beta_\nu = \mathbb{P}(\mathcal{M}_N < \gamma_\nu; \mathcal{H}_1) \leq e^{\gamma_\nu}$ , and the proof is complete. •

## REFERENCES

- [1] R. S. Blum and B. M. Sadler, "Energy efficient signal detection in sensor networks using ordered transmissions," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3229–3235, Jul. 2008.
- [2] Z.-Q. Luo, M. Gastpar, J. Liu, and A. Swami, "Distributed signal processing in sensor networks," *IEEE Signal Process. Mag.*, vol. 23, no. 4, pp. 14–15, Jul. 2006.
- [3] C. Chong and S. Kumar, "Sensor networks: Evolution, opportunities, and challenges," *Proc. IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug. 2003.
- [4] J. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, pp. 102–114, Aug. 2002.
- [5] J.-F. Chamberland and V. V. Veeravalli, "Decentralized detection in sensor networks," *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 407–416, Feb. 2003.
- [6] R. Viswanathan and P. K. Varshney, "Distributed detection with multiple sensors: Part I – fundamentals," *Proc. IEEE*, vol. 85, no. 1, pp. 54–63, Jan. 1997.
- [7] R. S. Blum, A. Kassam, and H. V. Poor, "Distributed detection with multiple sensors: Part II – advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64–79, Jan. 1997.
- [8] B. M. Sadler, "Fundamentals of energy-constrained sensor network systems," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, Aug. 2005.
- [9] S. Appadwedula, V. V. Veeravalli, and D. Jones, "Energy efficient detection in sensor networks," *IEEE J. Sel. Areas Commun.*, vol. 23, pp. 639–702, Apr. 2005.
- [10] C. Rago, P. Willett, and Y. Bar-Shalom, "Censoring sensors: a low-communication-rate scheme for distributed detection," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 32, no. 2, pp. 554–568, Apr. 1996.
- [11] W. P. Tay, J. N. Tsitsiklis, and M. Z. Win, "Asymptotic performance of a censoring sensor network," *IEEE Trans. Inf. Theory*, vol. 53, pp. 4191–4209, Nov. 2007.
- [12] S. Appadwedula, V. V. Veeravalli, and D. Jones, "Decentralized detection with censoring sensors," *IEEE Trans. Signal Process.*, vol. 56, pp. 1362–1373, Apr. 2008.
- [13] P. Addesso, S. Marano, and V. Matta, "Sequential sampling in sensor networks for detection with censoring nodes," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5497–5505, Nov. 2007.
- [14] P. Braca, S. Marano, V. Matta, and P. Willett, "Decentralized asymptotic detection by running consensus," in *Signal Processing Advances in Wireless Communications (SPAWC 2009)*, Perugia, Italy, June, 21–24 2009, pp. 41–45.
- [15] —, "Asymptotic optimality of running consensus in testing statistical hypotheses," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 814–825, 2010.
- [16] P. Braca, S. Marano, and V. Matta, "Running consensus in wireless sensor networks," in *Proc. of the Intern. Conf. on Information Fusion (FUSION 2008)*, Cologne, Germany, June 30 - July 3, 2008, pp. 152–157.
- [17] —, "Enforcing consensus while monitoring the environment in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3375–3380, 2008.
- [18] Y.-W. Hong, A. Scaglione, and P. K. Varshney, "A communication architecture for reaching consensus in decision for a large network," in *Proc. of the IEEE/SP 13th Workshop on Statistical Signal Processing*, July, 17–20, 2005, pp. 1220–1225.
- [19] E. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–115, Jan. 1954.
- [20] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Englewood Cliffs, N.J: Prentice-Hall, 1993.
- [21] H. V. Poor and O. Hadjiladis, *Quickest Detection*. Cambridge University Press, 2008.
- [22] P. Braca, S. Marano, V. Matta, and P. Willett, "Consensus-based Page's test in sensor networks," *Signal Processing*, vol. 91, no. 4, pp. 919–930, Apr. 2011.
- [23] S. A. Kassam, *Signal Detection in Non-Gaussian Noise*. Springer-Verlag, 1987.
- [24] B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics*. SIAM, 2008.
- [25] I. G. Bairamov and O. L. Gebizlioglu, "On the ordering of random vectors in a norm sense," *Journal of Applied Statistical Science*, vol. 6, no. 1, pp. 77–86, 1997.
- [26] B. C. Arnold, E. Castillo, and J. M. Sarabia, "Multivariate order statistics via multivariate concomitants," *Journal of Multivariate Analysis*, vol. 100, no. 5, pp. 946–951, May 2009.
- [27] R. Niu and P. K. Varshney, "Distributed detection and fusion in a large wireless sensor network of random size," *EURASIP Journal on Wireless Communications and Networking*, no. 4, pp. 462–472, 2005.
- [28] J. Galambos, *The Asymptotic Theory of Extreme Order Statistics*. Robert E. Krieger Publishing Company, 1987.
- [29] B. V. Gnedenko, "Sur la distribution limite du terme maximum d'une série aléatoire," *Ann. Math.*, vol. 44, pp. 423–453, 1943.
- [30] A. W. van der Vaart, *Asymptotic statistics*. New York: Cambridge University Press, 1998.
- [31] S. Marano, V. Matta, and P. Willett, "Distributed detection with censoring sensors under physical layer secrecy," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1976–1986, 2009.
- [32] S. Ross, *Stochastic Processes*, 2nd ed. New York: John Wiley & Sons, Inc., 1996.
- [33] S. I. Resnick, "Tail equivalence and its applications," *Journ. Appl. Prob.*, vol. 8, no. 1, pp. 136–156, 1971.
- [34] E. L. Lehmann, *Testing Statistical Hypotheses*. Springer, 2005.
- [35] H. L. Van Trees, *Detection, Estimation, and Modulation Theory. Part I*. New York: John Wiley & Sons, Inc., 1968 (reprinted, 2001).